DOCUMENT RESUME

ED 482 101 IR 058 789

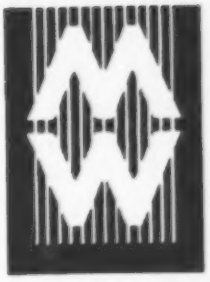
AUTHOR	Towler, Marjolein; Hobbs, Valerie; Pigott, Diarmuid
TITLE	Here and There: Managing Multiply-Purposed Digital Assets on the Duyfken Web Site.
PUB DATE	2002-04-00
NOTE	23p.; In: Museums and the Web 2002: Selected Papers from an International Conference (6th, Boston, MA, April 17-20, 2002); see IR 058 778. Some figures contain very small and illegible type.
AVAILABLE FROM	For full text: http://www.archimuse.com/mw2002/papers/towler/towler.html/.
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	Databases; Foreign Countries; Information Retrieval; Information Technology; *Metadata; Models; *Multilingual Materials; *Web Sites; *World Wide Web
IDENTIFIERS	Seventeenth Century; Netherlands; Ships; *Web Site Design

ABSTRACT

This paper describes a model for managing a dynamic Web site with multiple concurrent versions. The voyage of the Duyfken Replica (i.e., a replica of a 17th century Dutch ship) from Australia to the Netherlands has generated great interest in its associated Web site, and it is now planned to extend the original site to mirror sites in both Australia and the Netherlands, each completely bilingual, resulting in four different 'virtual' Web sites. However, the look and feel of the Web site needs to be the same across all the virtual sites, and the media resource used will be the same. This paper proposes a solution based on multiple sets of concurrent metadata and discusses some general implications of this approach for metadata harvesting and resource discovery. The prototype, which is based on a database that handles parallel sets of metadata, is described, together with the process for accession and annotation of media artifacts, and the dynamic generation of the Web pages from the database. The appendix contains the history of the Duyfken Web site. (Author/MES)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 482 101



Here And There: Managing Multiply-Purposed Digital Assets On The Duyfken Web Site

Marjolein Towler, CONSULTAS Pty Ltd., Valerie Hobbs and Diarmuid Pigott· Murdoch University, Australia

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

D. Bearman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Abstract

In this paper we describe a model for managing a dynamic Web site with multiple concurrent versions. The voyage of the Duyfken replica from Australia to the Netherlands has generated great interest in its associated Web site, and it is now planned to extend the original site to mirror sites in both Australia and the Netherlands, each completely bilingual, resulting in four different 'virtual' Web sites. However, the 'look and feel' of the Web site needs to be the same across all the virtual sites, and the media resource used will be the same. We propose a solution based on multiple sets of concurrent metadata, and discuss some general implications of this approach for metadata harvesting and resource discovery. We describe our prototype, which is based on a database that handles parallel sets of metadata, together with the process for accession and annotation of media artefacts, and the dynamic generation of the Web pages from the database.

Keywords: metadata, repurposing, narrative, dynamic Web site, media repository, bilingual

Introduction

The Duyfken Web site is the on-line exhibition of the Duyfken Replica. It serves to tell the story of the conception, construction and embarkation of the replica, and thereafter to recount its voyages around the world, much as the broadsheets informed the citizens of Haarlem and Amsterdam of the progress of the original voyage in the 17th Century. The Web site is the only exhibition available when the ship is between ports, and is currently charting the progress of the VOC2002 Voyagie. It is now planned to extend the original site to mirror sites in both Australia and the Netherlands, each completely bilingual, resulting in four different 'virtual' Web sites.

In this paper we describe the remodeling of the Duyfken Web site to accommodate multiple concurrent versions, of different languages and locations, dynamically delivered to the Web pages from a database and a media repository. The case of the Duyfken Web site provides us with a rare instance where we can examine the theoretical issues involved in multiple concurrency of semantic roles of media artefacts within the context of a practical example with real life pressures and constraints.

We first present a description of the Duyfken Replica project and the parallel development of the Web site. This demonstrates what the situation we are now facing is and how it came to be that way. We then describe the proposed remodelling of the site database and the revised procedures required to create it, drawing on the analysis of patterns in

IR058789

BEST COPY AVAILABLE

semantic usage and repurposing of media artefacts, and on the ways metadata is used to focus a resource to a goal-driven functionality.

The 1606 Duyfken Replica

In 1606 *Duyfken*, owned by the Dutch East India Company (VOC) and stationed in the East Indies, made a voyage of exploration looking for 'east and south lands' which took it on the first historically recorded voyage to Australia. As part of bringing Australian history to life, the *Duyfken* replica was built at the Lotteries *Duyfken* Village Shipyard in front of the Maritime Museum in Fremantle, Western Australia. Duyfken was built 'plank first', a method developed by Dutch shipwrights of the 16th and 17th Centuries. No original plans of any ship from the Age of Discovery exist because shipwrights did not use plans drawn on paper or parchment. The only plans were in the master-shipwright's head, and the ships themselves were built by eye. Although replicas or reconstructions of several Age of Discovery ships have been built in recent times, few of them seem to be able to sail anything like as well as the original ships, proving how little we understand of how these ships were built originally.

One of the stated objectives of the *Duyfken* Replica Project has been to produce a reconstruction that sails well enough to emulate the achievements of the original *Duyfken* (see http://www.duyfken.com/replica/experimental.html). So since she was launched, the finished Duyfken replica undertook the re-enactment expedition from Banda, Indonesia, to the Pennefather River in North Queensland, Australia, in 2000 (http://www.duyfken.com/expedition/index.html). She is indeed one of the few square rig replicas of that era capable of making an extended voyage under sail and, according to the known records, lives up to the performance of the original ship. She is currently sailing the original Spice Route in the wake of her 17th Century predecessor on her way to The Netherlands for the 2002 400 year VOC Celebrations (http://www.duyfken.com/voyagie/index.html).

Duyfken Web site - original

The Web site www.duyfken.com, built by Marjolein Towler and her team from Consultas Pty Ltd, is the on-line exhibition of the Duyfken Replica. It was initially developed to set the record straight. Four Duyfken Web sites had sprung up over a period of 6 months after the laying of the keel. None of these were under the control of the Replica Foundation and in some cases were simply factually inaccurate. The Web site was built to make correct information available about the original ship and its historical significance. Collaboration between developer (understanding technology and communication through digital media) and content expert (understanding on a deep level what is relevant to the topic)is crucial to any interactive media (Web) development, and Nick Burningham, maritime historian and the principal researcher and designer of the Duyfken replica, contributed his research and wrote all the original content.

The Web site also kept Web visitors informed of the building progress of the replica. The shipyard was open to visitors, so the physical building process was on exhibit; however, the Web site provided a central point of access to a wider audience, albeit only those with on-line access. The site was also used for Western Australian school curriculum activities, and it built a solid audience of followers from around the world.

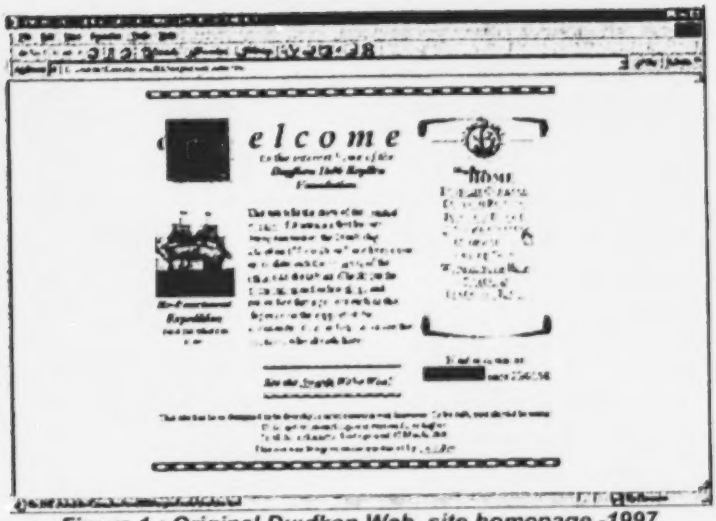


Figure 1 : Original Duyfken Web site homepage -1997 (detailed image)

The initial Web design was of a simple static HTML page nature. The structure took into account the kind of audiences we thought at the time would be interested in the topic: tall ship aficionados, primary school children, the Dutch and Dutch Australians, fellow maritime archaeologists, the Friends of the Duyfken association, historians, people interested in (model) ship building and ships in general, history, and wood working. This was a very broad audience to please, so although the information structure took the ship as its focal point, we tried to incorporate enough interest for any of the specific audience groups we identified.

The original Web site had two major topics: *Original Duyfken* and *Replica Duyfken*, which each expanded to a number of relevant subtopics (see Appendix). In addition there were general topics available such as Seafarers Links and a Sponsors page. Any visitor arriving at the homepage was presented with an introductory paragraph and could then traverse to either topic or all subtopics from every page.

To accommodate the different learning styles of users, we incorporated iconographic as well as text links. The screen layout was designed to fit on a 640x480pixel screen display and we limited the colour use to accommodate 256 colour displays. No moving images were added, since from a technology perspective we needed to accommodate low-end receivers and slow modem connections. We even limited the use of frames, because older Web browsers did not support them. The site was designed to look identical on both major browsers of the day (1997): Netscape and Internet Explorer, and we recommended version 3.0 or higher for both.

The only additions that were made to the Web site content in that time were the photographs of the replica building progress and after much agonizing - because it required a plug-in - QuickTime VR panoramas of the same topic. These were produced by Geoff Jagoe and Barb de la Hunty from Mastery Multimedia who offered their expertise. Collectively we argued that being able to see the ship through the panoramas did significantly add to the experience of the Web visitor, enough to justify

BEST COPY AVAILABLE

the wait for the plug-in download, anyway. For those users who did not want to go through with that, there were enough static images readily available.

Content changes and additions had to be made by the Web developers, because it required HTML editing. This was limited to once a month.

Figure 2 : Duyfken Homepage after Chevron 2000 Expedition was added - 2000 (detailed image)

duyfken.com – Chevron 2000 Re-enactment Expedition

After her launch, the ship took off on the Chevron 2000 Expedition. This re-enactment journey took her from Fremantle, Western Australia, to Jakarta, Indonesia. From there she traveled the original journey to the Pennefather River in North Queensland via the Spice Islands (http://www.duyfken.com/expedition/index.html).

This dramatically changed the requirement for the Web site, because the ship cannot be a physical exhibit when between ports. It thus became clear that the Web site is the *only* exhibit when the ship is at sea, and the only means of communicating her whereabouts to an audience. The kind of audiences interested in such a Web site expanded too. Adventure seekers, armchair travelers, sailing aficionados were added to the list. The site underwent a major overhaul in its layout and interface, partly to accommodate the Expedition section and partly to signify change, so regular visitors would realise that there was additional material.

The primary focus became the Expedition; the original content remained but became ancillary material. The navigation changed to reflect the different choices available and the Expedition section of the site was designed to look significantly different to indicate its separate status. However, we did make the clear choice to maintain the overall Web site homepage, and not make the Expedition homepage the main entry into the site. Our arguments were that the site was still the Duyfken Foundation Web site, not just the Expedition's, and that new visitors would not know of the entire topic choice if all they were presented with was the Expedition homepage.

BEST COPY AVAILABLE

We also made a concession to technology developments: the screen size grew to 800x600pixels and the colour settings to 16bit. After 3 years, nearly an eternity in technology terms, we felt that we would not leave too many users behind by doing this.

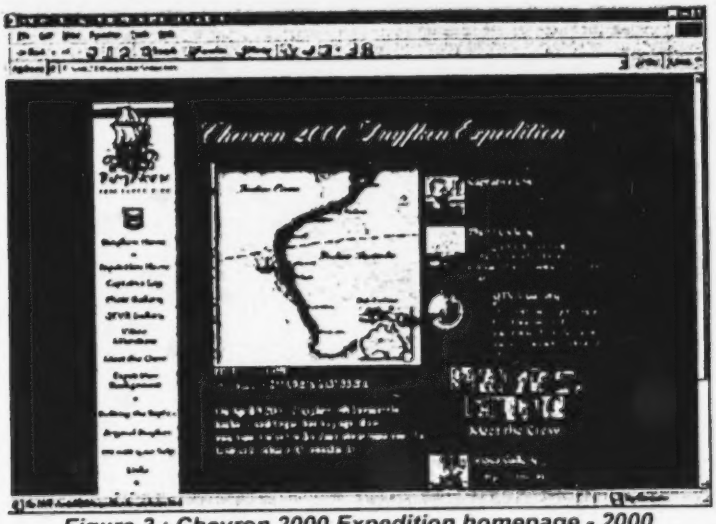


Figure 3 : Chevron 2000 Expedition homepage - 2000 (detailed image)

The content needs changed dramatically too. The project director Graeme Cocks wanted to explore the idea of incorporating the Captain's Log on the Web site with the potential to update it daily. This was the basis on which the dynamic database model was brought in. The captain of the ship at the time, Peter Manthorpe, e-mailed his logs via satellite to the land station, the Duyfken Foundation office in Fremantle. Anybody at hand there could and would go simply on-line to a secure URL and upload the logs into the database. Photo images of the journey were sent intermittently either when a reliable phone connection could be made, or when exposed film could be sent to Fremantle for development and scanning. Once digitized the images were uploaded in the database using the same procedure. This also happened to the QuickTime VR panoramas that Nick Burningham produced during the Expedition.

A bulletin board was added to the site to allow site visitors to leave messages. This system was set up to be moderated, but that was never explained well by the programmers, so it did not prove very successful. The project director used it a number of times for uploading press releases.

The database was built to receive the messages, but it was rather unyielding. The more Log messages were loaded, the longer it took to upload. In the end it was so slow the programmers of the system had to come in to change the 'time out' settings on the server. It was not user friendly to ordinary modem use, let alone slow connections in faraway places.

From a design perspective, there were limitations too. The database structure required page templates to be developed. Additional sections were not so easily incorporated, especially not if they required dynamic access, which needed specialist programming. It also required an ISP

BEST COPY AVAILABLE

that allowed client software to be placed on their servers, and that limited the choice of ISPs available.

The advantage was that anybody could upload text and image material through the administration console; it did not required any specialist knowledge of HTML or FTP. This was very important for the Foundation, since as a non-profit organization it was largely dependent on volunteers.

duyfken.com – VOC2002 Voyagie

Figure 5 : Homepage for Original Duyfken section – 2001 (detailed image)

Figure 6 : Homepage for the Building the Replica section - 2001 (detailed image)

BEST COPY AVAILABLE

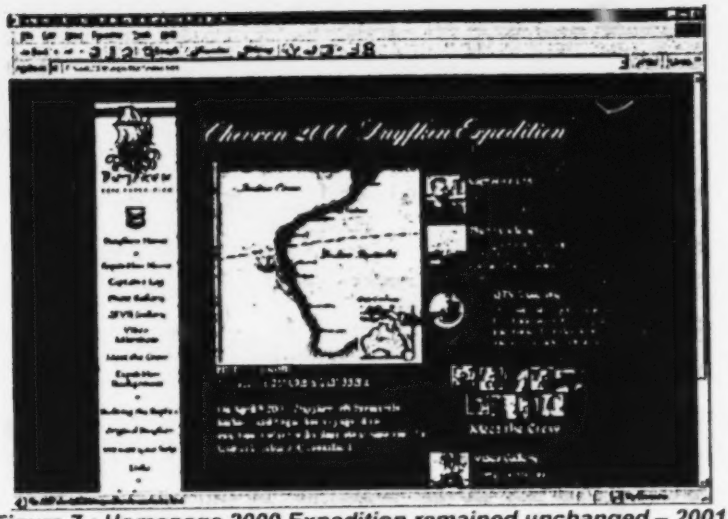


Figure 7 : Homepage 2000 Expedition remained unchanged – 2001 (detailed image)

Figure 8 : Homepage VOC2002 Voyagie section - 2001 (detailed image)

All the dynamic content in the Expedition database was transported to a static HTML format. This made that content more stable (although it increased the size of the Web site to around 600 pages). It also cleared the way to set up the new Voyagie section with a database connected through ASP, which can be edited dynamically through an HTML administration console on a secure URL.

The Voyagie section has five dynamic sections: Captain's Log, News & Events, Image Galleries, What's New and the Dutch version of the Captain's Log ('Scheepsjournaal'). Again, as with the Expedition section, the administrator can upload any data (text, images, QuickTime VR panoramas) through this console to those sections, including the daily logs the current captain, Glenn Williams, emails via satellite.

BEST COPY AVAILABLE

The database underwent a redesign from the original Expedition one. It was set up to allow for smaller monthly sections to be made. This limits the data entries to no more than 31 (the maximum days in a month), thus limiting the 'slowdown' effect too much data had on the upload procedure. It also not only has a secure URL, but now also requires password access to avoid any accidental or malicious access by unauthorized users.

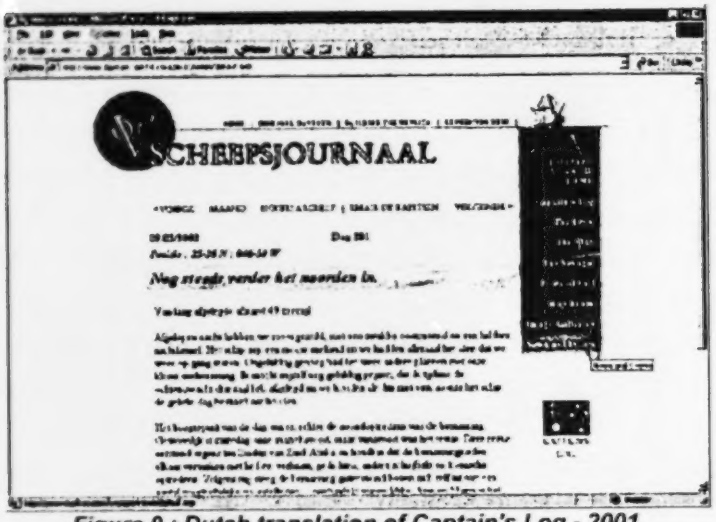


Figure 9 : Dutch translation of Captain's Log - 2001 (detailed image)

Figure 10 : Captain's Log main page - 2001 (detailed image)

BEST COPY AVAILABLE

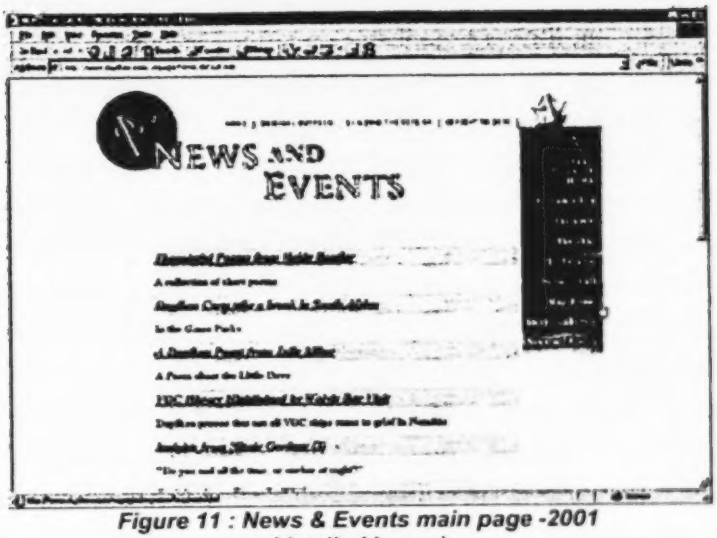


Figure 11 : News & Events main page -2001 (detailed image)

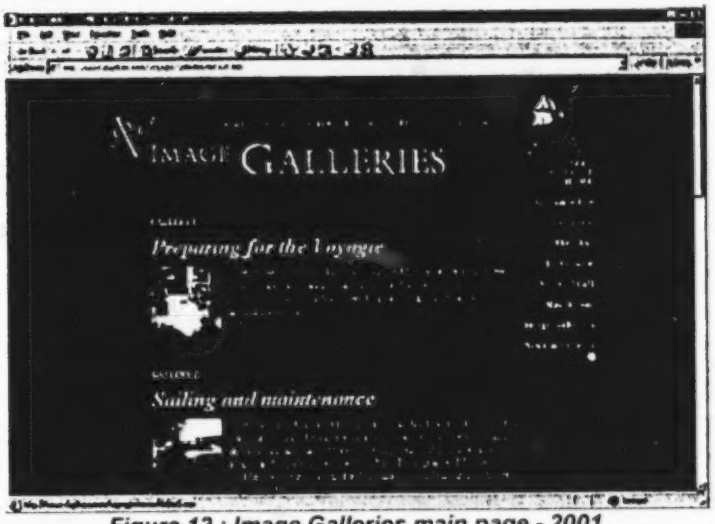


Figure 12 : Image Galleries main page - 2001 (detailed image)

The interface of the dynamic data pages also changed due to feedback from users. The Captain's Log pages of the Expedition were on a dark background, which was very dramatic and aesthetically pleasing as well as easy to read from the screen. However, most visitors wanted to print the logs out for keeping, or reading on paper, and we accommodated that by developing a light background for the Captain's Logs in the Voyagie section.

The VOC2002 Voyagie brought an even greater significance than was already there to the Dutch connection, and there was an extensive debate about making the site available in Dutch. At the moment the only Dutch part on the site is the 'Scheepsjournaal', the Dutch translation of the Captain's Log. The Duyfken is sailing to the Netherlands and will be traveling to all old VOC ports when she is there. There are great VOC

BEST COPY AVAILABLE

celebrations planned in the Netherlands throughout the coming year to commemorate the start of the VOC four hundred years ago. All this has posed some problems around control over the database, protocols of access to the database and the way the information is displayed to the individual Web visitor.

The Current Site and its Challenges

The Web site now contains of a mixture of archival, relatively static material retained from the original site, and the extremely dynamic Voyagie section that charts the progress of the replica day-by-day, providing an ongoing narrative of the expedition. Following the great interest generated in the voyage, the Duyfken Foundation wants to translate the entire site into Dutch, and in addition, for speed of access reasons, to create a mirror site of the entire development in the Netherlands.

The site currently consists of a single English language site sitting on an NT server at Web Central, an Australian ISP. The proposed changes effectively mean doubling the site and having both Dutch and English versions sitting side-by-side at Web Central, and a mirror of the site (Dutch/English) on an NT Server at an ISP in the Netherlands. The language issue dictates that there would be more than one person with access and in more than one location (Australia/Netherlands), and there could well be a need to have slightly differing information, especially in 'News & Events' once the Duyfken has arrived in the Netherlands. Thus, instead of a single site, we are looking at having four different 'virtual' sites, each potentially with differing content: English language in Australia, Dutch language in Australia, English language in the Netherlands, and Dutch language in the Netherlands.

All this has serious ramifications for the dynamic uploading of data through the administration console. With the division of the single site into four concurrent virtual sites, there is the potential for problems with control of content and media: it is essential that there be some sort of process control to ensure that one site does not lag behind the others, and that one site cannot be updated without the others, both in terms of copy (textual content) and media content (still images, movies, logs and so on).

This type of problem is normally addressed in software engineering by a version-control system, and for the Duyfken Web site we require a site-versioning system that permits multiple concurrency. What is needed is a system that not only enables the recording of files and their provenance and physical usage but also manages and preserves their contextuality and their semantic usage.

The problem of Multiply-purposed Media

While the revised Web site has at its core a transfer to a fully data-driven solution, its defining feature is that it is a metadata-driven solution. If a Website is to fit into the general milieu of RDF in the CIMI schema, the appropriate designation of semantic metadata to the media resource is essential (Dunn, 2000 ; Perkins, 2001) . On a static Web site, the metadata is assigned to the entire page, even if the significance of the metadata is the denoting of particular media resources within that page. In the context of dynamic delivery from a database and a media repository, such as we have here, the metadata must be preserved and

distributed at artefact level, if individual items are not to be bibliographically lost.

The critical moment in the life of a media artefact is when it is semantically bound, either directly to a purpose or by proxy to a purpose through a document such as a Web page. This is the only point at which intelligence is applied in the form of the skill and judgement of the curator. As media artefacts are at best only minimally self-referential, it is this act of binding that enables automatic sorting, selection and retrieval of sets of artefacts within the repository. The metadata derived from the use then stands for and represents the artefact within the system.

And yet, paradoxically, this act of representation is also necessarily a 'lens of distortion' (Lagoze, 2000) , and properly so - this is how we can focus on salient details at the expense of details of generalities. But this act of distortion means that in repurposing, the specific semantic role metadata cannot be reused, because there is no logical reason for assuming that this beneficial distorting of reality will be equally beneficial to the role anticipated in the repurposing. In other words, although the usages are themselves pertaining to the same artefact, they represent different aspects of reality.

Moreover, every subsequent reuse modifies the net effect of the pre-existing usages. We have argued elsewhere that there must be a set of protocols established in repurposing to ensure that meanings stored in multimedia databases do not 'drift' (Pigott, Hobbs, & Gammack, 2001) .

There are two aspects of multimedia repositories that have implications for metadata practice and artefact repurposing. At system level, we can see that the repository has a telos, or purpose, and intended audience, within its overall theme: adult and child versions of a museum site, or subscriber and free versions of a digital library; or here, the four virtual sites of the Duyfken project. At artefact level, we can consider the usage pattern that the artefact follows throughout its life, through use, repurposing and archiving, which has implications for the extent to which usage is generic or specific, or lying at some stage in between.

These twin dimensions require that we have two aspects to the metadata also: the *teleological* to target the intended audience, and the *abstractive* to ensure the correct level of specificity is used.

When we consider the teleological aspect, we can see that for each instance of metadata usage, there must be a prescribed context of metadata usage, which gives explicit instructions on what constraints are being used, and how these rules are to be applied. This we can term the *metadata frame of reference*, as that is what gives a meaning to the individual usage. A proper program of metadata usage therefore involves the objects for annotation, the instances of annotation, and the rules and keyword sets (the authority set) to use for that annotation.

When we consider the usage pattern aspect, we see that when a media artefact first comes into contact with the repository, it is the decision of the curator that determines its usage pattern, and hence the specific or general nature of the metadata required. There are three main patterns:

- The *immediate permanent use* pattern is where an artefact arrives and is immediately put to use. In this case, it must be catalogued with metadata specific to that use.

- The *immediate archive* pattern is where there is no immediate use for the artefact, but it still has the potential for some later, unspecified usage. Here the metadata needs to be general to enable the artefact to be retrieved for a variety of purposes, none of which is known at the moment of accession.
- The *immediate transitory use* pattern is where the artefact is to be used for some short-term specific purpose and then archived. Here the curator needs to strike a balance between the short-term specific and long-term generic metadata.

There is a special case of usage pattern that we meet in the Duyfken project: the *immediate concurrent use* pattern, where an artefact has more than one semantic purpose *at the moment of accession* (permanent or transitory). These parallel meanings will remain linked through at least the initial part of the artefact's life. Here we need to have multiple sets of specific terms, one for each purpose, and to be able to manage and exploit that relatedness within the overall systematic of organization and retrieval.

When we superimpose the usage pattern of the artefact on top of its teleological aspect, we meet with the problem of continuity and change described in Pigott et al. (2001) , where we discussed the survival of authorial intent; and in Gammack, Pigott, & Hobbs (2001) . If early binding to a specific use requires specific metadata, any later repurposing to another specific use will require metadata specific to that use, a situation which is at least potentially conflicting. If the artefacts have been catalogued with generic metadata, however, a later specific use can be accommodated. This is exactly the problem that we must manage in establishing the systematics of the virtual sites of the Duyfken Web site.

Parallel Metadata: a Prototype for the Revised Duyfken Web site

Our solution is built on a database that lets us represent concurrent use in the Web site, in the form of parallel metadata for media artefacts stored and maintained for the four virtual sites, so that the correct version is available to be called upon for a particular site and language. Over and above this base requirement, we need to ensure that we represent the components in such a way that the site of which they are a part can preserve its narrative structure, in order that we may call on those resources in the correct context.

We therefore have several clear objectives for the prototype:

1. It must retain the existing style of the site, and permit it to expand to four separate, yet parallel, sites.
2. To accomplish this, it must permit separate sets of metadata for each of the four concurrent usages (English language–Australian site, English–NL, Dutch–Australia, Dutch–NL).
3. This metadata should fit into the industry standard practices of defining and exploiting metadata in order that the local expenditure of effort have a universal return.
4. There should be a naturalistic process for the accession/cataloguing of media, to permit an unobtrusive capturing of the metadata and its storage for the appropriate virtual site.
5. There should be a process for generating the Web pages in

context that gives rise to the dynamic delivery of the four separate virtual sites, a process which should also ensure the 'narrative' of the particular site is preserved.

When we modeled the database, we identified four main entities of interest: Event, Setting, Place and Agent. The data model centres on the concept of an EVENT. Events are things that happen; on the voyage, or before or after it – stages in building the replica, the departure and arrival of the ship, individual passages and stays at ports, and the news and events, including captain's log, that are mailed regularly to the Web site. Events are linked with other events through a recursive relationship, as an event such as a stay at a port is also part of the larger event of the relevant passage, which itself is part of the entire voyage. We also note that since the Web site is itself part of the events of the voyage, charting as it does its chronology and narrrative, the construction of each Web page will also be recorded in the database as an event, and the Web pages themselves stored as artefacts.

Events occur at Places, such as ports or unnamed locations of latitude/longitude; and in a particular Setting, which could be on the ship, a particular part of the ship, or onshore. Events have Agents associated with them. Agents can be individual people, crew positions, or collective groups such as crew or land staff. Again, each of these entities is in a recursive relationship with itself, as each place or setting may be a part of a larger whole, and individual agents may be part of a group.

The relationships identified among the entities enable the narrative structure of the Voyagie site to be represented, as well as the more static information present in the original site. Since all the main entities are in recursive relationships, the inherent hierarchies in the Duyfken voyage can be modeled. This allows us to extract more complex combinations of information, for example getting connections between legs of the journey and members of crew, via ports of call, and the watches of the night.

In the Duyfken Web site, we have four virtual sites, consisting of the combinations of two languages and two countries. So, for each physical media artefact playing its particular role in the narrative, we need to consider four sets of metadata. We can identify several possibilities resulting from this:

- The language is different, but the site makes no difference to the metadata that is used. This occurs with the technical metadata, where there is no interpretation involved, and also with the conventional data in the Event, Place, Setting and Agent tables. This situation can be accommodated by having dual language fields within the same record. (Although this is a 'hard coding' solution it is simplest to manage and is justifiable here as it is unlikely that any other languages will be required for the site.)
- The site at which the artefact is used may impose a different interpretation on it. For example, the same image could be captioned differently in the Netherlands and in Australia. This involves not just language translation, but a different set of semantic metadata. For this situation, we use four parallel metadata records.
- If the interpretation or translation required for a site involves editing or copying the artefact, we end up with multiple physical artefacts: for example, each entry in the Captain's Log/Scheepsjournal is stored as two separate documents. Each

document, being a separate artefact, would then require dual language fields for the technical metadata and four parallel records for the semantic metadata, as described above.

Process for Accessioning

The Web site, and therefore the database which underpins it, is a live chronicle of events. The selection of media for a section of the narrative is done by a series of page wizards, permitting searching for material either generally or specifically, and from the perspective of an entity or from a unified set based on the metadata. An interesting challenge here is to present multiple possibilities for keyword or character/place referencing without losing the flow of the selection process.

The Web pages are designed to have zones of placement, as close to the current Duyfken site as possible. Use of a cascading style sheet in conjunction with the stored text permits a set of different semantically purposed page regions, each with content that is apposite and current. The use of summary textual metadata permits 'headlines' to be created, and serve as keyholes to the pages where the information is given in full, and also from each section to a set of previous entries for that section. This fits in nicely with the image of the site as a narrative chronology, similar to that in old Amsterdam. There is a strong publishing feel to the site, which enhances its feeling of vitality and immediacy.

The new Web site will have curators in each country whose responsibility it is to add the media and data to the database. Although there need to be four 'virtual curators', one for each site, the roles could be filled by anything from one to four or more individuals; thus in the following discussion, 'curator' refers to the appropriate virtual curator. The secure access of the existing site will be continued, and curators will be required to log in with passwords.

Media artefacts and notification of events will arrive regularly, as the voyage progresses, and the database, media repository and Web pages will be updated at the same time, in the context of the ongoing narrative. Figure 13 illustrates the logical process of accession of media and data and the delivery of Web pages.

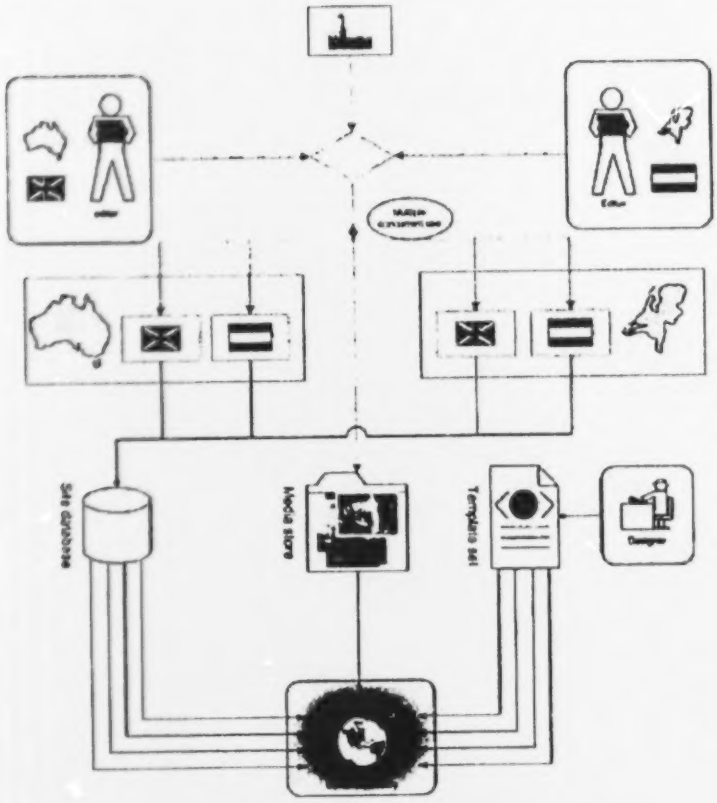


Figure 13: Process for accessioning (detailed image)

The new site is 'event driven', and there are three different kinds of stimulus from the world that the curator must respond to:

- The arrival of new media artefacts, such as images or an e-mail of the Captain's Log. These artefacts will always be part of the context of an event, whether new or previous.
- An event happening, such as arrival at a port (which may or may not have accompanying media).
- A general need for updating the site.

Each of these stimuli is a trigger that causes the curator to be notified, using internal messaging, that the system requires attention. We will now describe each of these processes in turn.

When new media artefacts arrive, there is likely to be a surplus of artefacts, and the first judgement of the curator will be one of appropriateness of the artefact (be it still image, recording, e-mail, movie) to the immediate site narrative. In some instances (the Captain's Log is the main one) the artefact is immediately needed for all locations (and in the case of the log, it must be translated and a separate document created). With a set of volunteer images of a stay at a port, on the other hand, it may be that only three images out of 20 would be used.

BEST COPY AVAILABLE

The next question arises as to the usage status and the metadata requirements of the remainder. As discussed previously, there will be a tradeoff between very specific metadata for immediate use, and more general metadata for archival use. The artefacts that are immediately archived are often going to be very similar to the ones selected for immediate use, and a level of description and keywording will be aimed for to permit a general location, but perhaps not the detailed analysis of action and character of an artefact designated for immediate use.

It is essential that every artefact receive its full complement of parallel metadata, and that this process be managed. A system of locking will ensure that it is impossible for more than one curator simultaneously to access and catalogue the same artefact, and a set of pages inserted after the initial accession pages present in the original facilitates the addition of the metadata. The metadata will be drawn from best practice technical metadata requirement and the selected semantic role metadata sets.

The first curator to respond to the message would be the one to add metadata from an agreed set, with 'place-holding' metadata put in the system for subsequent curators. The place-holding metadata will be generic rather than specific, and will be in the appropriate language for the usage, drawn from key-paired thesaurus and similar wordsets, wherein every entry in the thesaurus will have a Dutch as well as an English term. The titles, descriptions and other free text will have to be left with a translator, or team of translators, who will be delegated the individual tasks automatically by internal messaging. The next point in the chain is a form of 'copying-on', whereby the archived (not immediately used) artefacts are given some of the used, generic, metadata automatically (Figure 14).

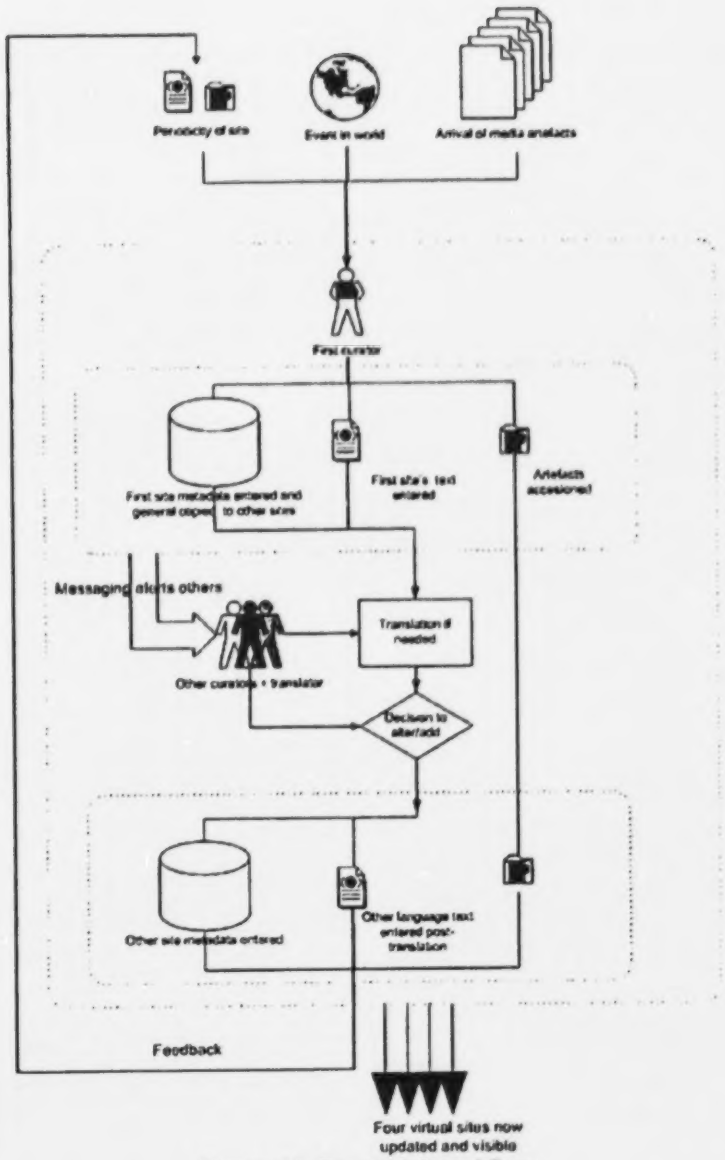


Figure 14. Curatorial workflow. (detailed image)

The second curator on the scene follows the same process of selecting from the artefact set and accepting or amending the default metadata, and if necessary adding specific metadata. It may also be that the second curator finds an alternative artefact in the batch to suit their purposes better, in which case, the two sites will not have identical media artefact usage.

The artefacts are all placed into a media repository, which is automatically mirrored, and updated daily. This ensures that there are no bandwidth constraints in the delivery of media for analysis.

As the media artefacts will always added be in a context – the context of

BEST COPY AVAILABLE

an event, agent, place or setting – the possibility of adding a new record to the database tables at the same time has to be taken into account. Again, the first curator would enter the information, and the other language version would be delegated to the translators.

The second case that the curator must respond to is when events occur and trigger an update to the database but without accompanying media. In this case, the curator may need to search through the existing media repository to locate suitable media artefacts for illustration.

Finally, there is always the need for the Web site to be fresh and interesting for visitors, so in the absence of any new media or events to be recorded, it may be necessary for the curator to write copy on more general themes. This could be prompted on a time or calendar basis. Here the curator may need to search the database for suitable media and perhaps also for 'stories' from the records of events.

After the completion of each update of a Web page at a virtual site, the creation of the page is itself recorded as an event and the Web page is stored as part of the media repository. An interesting side-effect of the Web site creating dynamic pages as a chronicle is that the contents, while dynamic, are representative of their period of currency, and so themselves become the subject of harvesting and storage. It is interesting to consider whether the individual pages will require the direct attention of the editor for adding metadata – presumably the requirement for internal monitoring of the site could well have this as a clear benefit, with the response of the site to certain events being indicated with candour for internal purposes where they would not have been evident from the site itself.

What distinguishes the Duyfken project is the need for the parallel versions to remain concurrent in time throughout the constantly dynamic process of updating the narrative. Ideally, updates to the four virtual sites would be handled as a single transaction, so that no version lags behind the others. In practice, however, it may be necessary to set a timeout period so that if for any reason a curator fails to respond to a message by posting an update, the remaining sites are not held up indefinitely.

Implementation

The Web site is currently on a Windows NT server using ASP, with a perl/mysql backend delivering media as required. There is a hybrid ECMAScript (JavaScript) and ASP development environment, showing the different stages of the project development, as described in the introduction. The initial prototype, replacing the Voyagie/ship's journal section, was done with the Pasigraphy scripting environment (developed by D. Pigott) and had the data in Microsoft Access tables under IIS5.

The next stage, the pilot project, makes use of a straightforward ASP/ADO format, designed and managed through Microsoft FrontPage. The interoperability of IIS 5 with FrontPage and the relative ease of use for the non-technical user will make it possible for the curators to experiment with alternative features in relative safety, due to the componentisation of the design. The choice of this platform was made partly for reasons of continuity, and partly with a view to making a generalisable solution that would be usable elsewhere. The tables are currently in a SQLServer7 backend, but upgrading to SQLServer 2000 presents no challenges at all.

The movement of files is done via HTTP Upload to avoid any potential confusion associated with standard UNIX ftp. The exchange of metadata and the new entities is done automatically by time-triggered stored procedures, while the inter-curatorial communication is done by an internal, project-style, messaging system. It is hoped that by designing the system to be scalable, there will be no barrier to repurposing the system itself.

The design of the database was done using the entity-media modeling methodology (Hobbs & Pigott, 2002) and the process of designing for concurrency is described in Hobbs, Pigott & Towler (in prep). An unusual feature of the design is the recursive nature of the four principal entities involved, which has led to an unusually adaptable system for querying and producing network reports. The parallel metadata as required for the virtual sites is stored using dual language fields for technical metadata and conventional data, and separate records for semantic metadata. Keyword sets with linked terms are used to accommodate the complex nature of bilingual synonymy.

In terms of the choice of media artefact formats, although the Duyfken project is to a large degree archival, there has had to be a trade-off between the needs of archiving material (long term, guaranteed readability, multiple store, etc) and the needs of the site, which is to be maximally viewable, with an optimum download. Certainly the Duyfken voyage is of interest to places where it calls on its way, and there is a responsibility of the site to permit easy viewing for those on the other side of the digital divide. Moreover, there has been no way of guaranteeing a single standard for quality where some of the digital material is volunteered, and already arrives in a digital format. The choice to date has been to opt for high quality JPEG formats, QuickTime VR and HTML4 compliant text. With the transition to the new site, the needs of longevity are probably better served by continuity, due to the massive task of retroconversion. Certainly, embracing formats such as fractal compression or wavelet compression ahead of their general availability in browsers would go against the Web site's principle of universality of access.

The Duyfken Web Site and Resource Discovery

The late-binding use of media and text resources in the revised Web site not only makes for timeliness and accuracy, but also makes it possible to have an inventory of the entire media resource pool, and the benefits that entails. Not only are there savings in time and money through the reuse of media artefacts, but we are also moving towards an environment of repurposing and media discovery, and the increased semantic richness that arrives with that discovery. It is the intention that (though the pages are dynamically created) the Duyfken Web site be a source for Dublin Core compliant harvesting, to permit discovery by other museum and historical presences on the Internet. But in facing up to the challenges this presents, some problems are immediately apparent.

The IFLA FRBR common logical method for metadata presents the information resource in one of four states: Work, Expression, Manifestation and Item (Bearman, Miller, Rust, Trant, & Weibel, 1999) . It s not straightforward to fit the work of the Duyfken Web site into this paradigm, however, for several reasons.

The fundamental nature of the DC paradigm is of works of cultural

significance being described and accessed, reflecting its bookish and artistic heritage. This is not readily applicable to our situation. The artefacts are not necessarily of worth in themselves, and there is a documentary rather than an artistic presence on the Web site (this is not to deny the aesthetic worth, but rather to describe its direction). There is no real creator or authorship of importance, but rather the *story* itself is the main bearer. The sailors themselves are telling the story through their actions, and that is the authorial component that should be stressed.

Neither is there any manifestation of importance. The presence of the media artefacts here are all of an illustrative nature: here we are concerned with trying to represent events, and artefacts are chosen for their ability to represent that narrative, rather than for any significance in themselves.

That said, what can we gain from the insight of the OAG and the DC? Certainly the paradigm of interoperability is crucial here: permitting the discovery of information through judicious use of RDF, and through the appropriate access to properly accessioned media artefacts in a repository, is the responsibility of the conscientious cultural store point-of-presence – so what can be asked of the site itself in this paradigm? Here we find resource discovery not so much between the harvester and the site components, but as part of the harvesting of the site itself as a resource; so metadata practice should have as its first priority harvesting of the site as a narrative rather than individual points of access to the site.

Perhaps it would be more appropriate to consider the types of requests that might be made of the repository in actual use. Here questions would be asked of the site as of an expert on Dutch maritime history, so the resources could be exposed and harvested in the form of specific rather than general queries. Not so much 'what media have you available' as 'what can you tell me about this subject in particular?' This goes back to our earlier discussion of how metadata information should be prepared for early and late binding use.

References

Bearman, D., Miller, E., Rust, G., Trant, J., & Weibel, S. (1999). A Common Model to Support Interoperable Metadata: Progress report on reconciling metadata requirements from the Dublin Core and INDECS/DOI Communities. *D-Lib Magazine, 5*(1).

Dunn, H. (2000). Collection Level Description - the Museum Perspective. *D-Lib Magazine, 6*(9).

Gammack, J. G., Pigott, D. J., & Hobbs, V. J. (2001). *Context as history: the cat's cradle network*. Paper presented at the Australasian Conference on Knowledge Management and Intelligent Decision Support - ACKMIDS '2001.

Hobbs, V. J., & Pigott, D. J. (2002). Entity-media modelling: conceptual modelling for multimedia database design. In D. R. G. Harindranath, John A.A. Sillince, Wita Wojtkowski, W. Gregory Wojtkowski, Stanislaw Wrycza and Joze Zupancic (Ed.), *New Perspectives on Information Systems Development: Theory, Methods and Practice*: Kluwer Academic, New York, USA.

Lagoze, C. (2000). Metadata Musings, *Talk presented at University of Virginia, April 2000.*

Perkins, J. (2001). *A new way of making cultural information resources visible on the Web: Museums and the Open Archives Initiative.* Paper presented at the Museums and the Web 2001, Seattle.

Pigott, D. J., Hobbs, V. J., & Gammack, J. G. (2001). An approach to managing repurposing of digitised knowledge assets. *Australian Journal of Information Systems* (Special Issue on Knowledge Management).

Appendix: History of the Duyfken Web site

Web site Credits:

Nick Burningham: Maritime Historian/Archaeologist – Web site Content Writer|Researcher – **nickb@iinet.net.au**

Scott Ludlam: Graphic Interface Designer|Web Authorer

Geoff Jagoe & Barb de la Hunty: QTVR panorarna specialists – **geoffandbarb@mastery.com.au**

Marjolein Towler: Project Manager|Information Designer|Artistic Director – mt@consultas.com.au

Feb97	**The building of the replica starts** Four unofficial Duyfken Web sites spring up in a very short time. The Foundation has NO control over any of them, although it gives cooperation to some. Unfortunately, they contain information the Foundation regards as incorrect or misguiding.
Sep97	Consultas is approached to produce the official Foundation site.
	The initial structural and information design are developed in close collaboration between Marjolein Towler (Consultas Project Manager, Creative Director) and Nick Burningham (Foundation's historian/archaeologist and ship-designer) as the content expert.
Jan98	The site is launched with the following structure: • **Original Duyfken** • **Replica Duyfken** (This was the dynamic section – changing monthly) • Sponsors • Seafarers' links • Competition • Shop • We need your help
Jan98 – Jan99	**Growth of the *Replica Duyfken* Section (maintenance by Web developer)** Monthly updates to photo-archive of the building progress.
Apr98 – Jan99	Additions of QTVR panoramas - Mastery Multimedia, experts in creation of QTVR panoramas offered to provide panoramas of the building process
Jan99	**Duyfken is launched** Short MPEG videos of the ceremony are posted on the site within 1 hour.
1999	During rigging, fitting out and sailing trials, QTVRs and photo-archive updates continue.
Apr00	**The Chevron 2000 Duyfken Expedition: The replica departs Fremantle on a long re-enactment journey to Queensland, Australia, via Banda, Indonesia.** The site undergoes a major redesign (new interface) and expansion incorporating the existing site's structure and content as a static (unchanging) archive:
	New Expedition Section (maintenance/content control by Web site owner): *Dynamic Features:* - Weekly updated map on expedition homepage, Daily up-dated 'Captain's Log', Expedition photo-archive, Expedition QTVR archive and Bulletin board. *Static Features (maintenance/control by Web developer):* - Background information on voyage: geography, politics, climate, history, etc. and Crew profiles All dynamic content is now in a database, which is controlled by the Web site owner (the Foundation) via a secure URL through ASP management console. This gives controlled facility in uploading new content, but reduced flexibility in structural redesign.
May01	**The VOC2002 Voyage: The replica departs Sydney on a long re-enactment voyage from Jakarta, Indonesia to Texel, the Netherlands via the original spice route.** The site undergoes again a major redesign (new interface). The 2000 Expedition material is migrated from the database to static HTML and the VOC2002 section is added. The homepage is changed to incorporate a dynamic element, which is controlled from the secure URL administration console.
	New VOC2002 Voyage Section (maintenance/content control by Web site owner) *Dynamic Features:* - Daily up-dated 'Captain's Log' (and its Dutch translation: Scheepsjournaal), Image Galleries, News & Events *Static Features (maintenance/control by Web developer):* - Background information on voyage: history, culture, etc., crew profiles, the map room, ports of call.
Nov01	Collaboration commences between Consultas and V. Hobbs and D. Pigott to investigate redesign of database to accommodate different virtual Web sites.

Table One: Site Chronology (detailed image)

BEST COPY AVAILABLE

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

[X] This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

[ ] This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").